Which Moments to Match?

Gallant and Tauchen (1996, Econometric Theory)

Bálint Szőke

February 9, 2016

New York University

Introduction

Estimation

"Usual" setting:

- We want to characterize some data: $\{\widetilde{y}_t, \widetilde{x}_{t-1}\}_{t=1}^{T}$
- Assume a family of parametric models (= likelihoods)

$$\{p(y, x|\rho) : \rho \in \mathcal{R}\}$$

  where the true DGP is $\rho_0 \in \mathcal{R}$
- p is called the structural model
- For simplicity, let p be strictly stationary and ergodic
- Aim is to find (estimate) ρ_0

Alternative approaches

1. If feasible we can use Maximum Likelihood:

$$\widetilde{\rho}_0 \equiv \text{argmax}_{\rho \in \mathcal{R}} \sum_{t=1}^{T} \log p(\widetilde{y}_t | \widetilde{x}_{t-1}; \rho)$$

2. If infeasible or you want to *"do something without having to do everything simultaneously"* (LPH) we can use GMM
   - papers by Burnside, Christiano and Eichenbaum
   - moments are "arbitrarily" picked (like calibration)

Key idea: MLE $\approx$ "GMM on the scores"

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \rho} \log p(\widetilde{y}_t | \widetilde{x}_{t-1}; \rho) = 0 \qquad \mathbb{E}^{p(y|x;\rho_0)} \left[\frac{\partial}{\partial \rho} \log p(y|x; \rho_0) \right] = 0$$

MLE finds the statistically most informative moments

Gallant and Tauchen (1996)

Efficient Method of Moments

New object: auxiliary model or score generator, $f(x, y|\theta)$

- (1) fits the data well, (2) tractable (can calculate scores)
- Not necessarily "structural", e.g. VAR, ARCH, GARCH
- like a reduced form equation in classical SEM

Idea: if f fits the data that is generated by the structural model p, i.e. $f \approx p$, we can replace the score of p with the score of f in

$$\mathbb{E}^{p(y|x;\rho_0)}\left[\frac{\partial}{\partial\rho}\log p(y|x;\rho_0)\right] = 0 \quad \text{so} \quad \mathbb{E}^{p(y|x,\rho)}\left[\frac{\partial}{\partial\theta}\log f(y|x;\theta)\right]$$

should be (close to) zero at the true ρ_0, if p is correctly specified

- Use this vector of moment conditions for GMM
- How to get rid of θ? How do we calculate expectations under p?

Efficient Method of Moments – Algorithm

Step 1: *projection* (data onto the auxiliary model)

- f is fitted by ML to get $\tilde{\theta}_T$. Then the score over the data satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta}\log f(\widetilde{y}_t|\widetilde{x}_{t-1};\tilde{\theta}_T) = 0$$

- Calculate

$$\widetilde{\mathcal{I}}_T = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial}{\partial\theta}\log f(\widetilde{y}_t|\widetilde{x}_{t-1};\tilde{\theta}_T)\right]\left[\frac{\partial}{\partial\theta}\log f(\widetilde{y}_t|\widetilde{x}_{t-1};\tilde{\theta}_T)\right]^T$$

Efficient Method of Moments – Algorithm

Step 2: *estimation*

- Suppose that we can simulate p for given ρ to get $\{\hat{y}_n^{(\rho)}, \hat{x}_{n-1}^{(\rho)}\}_{n=1}^N$. For large enough N, define

$$m(\rho, \theta) \equiv \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial \theta} \log f\left(\hat{y}_n^{(\rho)} | \hat{x}_{n-1}^{(\rho)}; \theta \right) \approx \mathbb{E}^{\rho} \left[\frac{\partial}{\partial \theta} \log f\left(\hat{y}_n^{(\rho)} | \hat{x}_{n-1}^{(\rho)}; \theta \right) \right]$$

- If the DGP is $p(y|x, \rho_0)$, we expect $m(\rho_0, \tilde{\theta}_T) = 0$
- Estimator for ρ_0 is

$$\hat{\rho}_T \equiv \text{argmin}_{\rho \in \mathcal{R}} \; m^T(\rho, \tilde{\theta}_T)(\widetilde{\mathcal{I}}_T)^{-1} m(\rho, \tilde{\theta}_T)$$

Asymptotic properties

For fixed f and p, let θ_0 be defined so that $m(\rho_0, \theta) = 0$ is satisfied.

Asymptotics

Under standard regularity conditions

$$\lim_{T\to\infty} \hat{\rho}_T = \rho_0 \quad \text{a.s.}$$

$$\sqrt{T}(\hat{\rho}_T - \rho_0) \Rightarrow \mathcal{N}\left(0, [(M^0)^T(\mathcal{I}^0)^{-1}M^0]^{-1}\right)$$

$$\lim_{T\to\infty} \hat{M}_T = M^0 \quad \text{a.s.} \quad \text{and} \quad \lim_{T\to\infty} \widetilde{\mathcal{I}}_T = \mathcal{I}^0 \quad \text{a.s.}$$

where $M(\rho, \theta) = \frac{\partial}{\partial\rho} m(\rho, \theta)$, $M^0 = M(\rho_0, \theta_0)$ and $\hat{M}_T = M(\hat{\rho}_T, \tilde{\theta}_T)$.

If for some open neighborhood R_0 of ρ_o, there is a twice continuously differentiable $g : R_0 \mapsto \Theta$, s.t. $p(y|x, \rho) = f(y|x, g(\rho))$, then the estimator has the same asymptotic distribution as the MLE of the structural model.

Score generator – guidelines

- Score in closed form AND closely approximates the data
- Identification
  - Order condition: dim(θ) $\geq$ dim(ρ) (minimal complexity)
  - Rank condition: nonlinear framework, hard to check. In case of flat spots, we can adjust the score generator to include higher moments.
- Example: for financial data, sequence of densities defined by an ARCH or GARCH process
- General purpose score generator: **Seminonparameteric (SNP)**
  - K-truncated Hermite exp of the square root of an innovation density
  - expected to closely approximate any nonlinear Markovian process
  - tractable and flexible (easy to adjust to certain features of the data)

Isn’t it the same as indirect inference?

Indirect inference (Wald approach)

Three main steps

Step 1: MLE of $\log f$ on the data to get $\hat{\theta}$

$$\hat{\theta}_T \equiv \text{argmax}_\theta \quad \log f(\widetilde{y}, \widetilde{x}|\theta)$$

Step 2: MLE of $\mathbb{E}^p[\log f]$ to get the **binding function** $\theta(\rho)$

$$\theta(\rho) \equiv \text{argmax}_\theta \;\; \mathbb{E}^{p(y,x|\rho)}\left[\log f(y^{(\rho)}, x^{(\rho)}|\theta)\right]$$

Step 3: Estimate ρ_0 with

$$\hat{\rho}_T \equiv \text{argmin}_{\rho\in\mathcal{R}} \;\; (\hat{\theta}_T - \theta(\rho))^T W(\hat{\theta}_T - \theta(\rho))$$

where $W = \left[\mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-1}\right]^{-1}$ $\quad \mathcal{I} = \mathbb{E}^{p(y,x|\rho)}\left[\left(\frac{\partial}{\partial\theta}\log f(y,x|\theta(\rho))\right)^2\right]$

and $\mathcal{J} = \mathbb{E}^{p(y,x|\rho)}\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(y,x|\theta(\rho))\right]$

So are the two the same?

YES

- both approaches use auxiliary model as an adjunct
- for any given auxiliary model same asymptotic distribution

NO

- EMM computationally less intensive (for the binding function we reestimate the auxiliary model for each ρ), espec. with non-linearities
- *" The auxiliary model does not need to be an accurate description of the data generating process. Instead, the auxiliary model serves as a window through which to view both the actual, observed data and the simulated data generated by the economic model: it selects aspects of the data upon which to focus the analysis."* (A. Smith) $\approx$ GMM with "arbitrarily" picked moments

References

- Gallant and Tauchen (1996) – Which Moments to Match? *Econometrics Theory*
- Gallant and Tauchen (2010) – Simulated Score Methods and Indirect Inference for Continuous-time Models. *In: Handbook of Financial Econometrics Edited by: Yacint Ait-Sahalia and Lars Peter Hansen*
- Anthony A. Smith Jr. (2008): Indirect Inference. *In: The New Palgrave Dictionary of Economics, Second Edition*